

Summary

Contents

CONTENTS

**Data**

Includes: **Definitions, Example**
Data are observations or facts that can become information or knowledge

**Index**

Includes: **Definition:, What do indexes tell you?, How can we calculate an Index?, When is an index not appropriate?**
An Index is a number used to show the variation in some quantity over time.

**Mean**

Includes: **Definition , Here's another way of looking at it!, How do you calculate the mean?, When is the Mean not a useful measure?**
The mean is a summary number that measures the midpoint of a dataset.

**Measures of Error and Spread**

Includes: **Definitions, Calculations, Using Standard Deviation, Standard Error and Relative Standard Error**
Standard deviation measures the scatter in a group of observations

**Median**

Includes: **Definition, What can the Median tell you?, How do you calculate a Median?, When is the Median not a good measure of central tendency?**
The median is the middle score that separates the lower half of a dataset from the higher half.

**Mode**

Includes: **Definition , What does the mode tell you?, How do you calculate a mode?, When is the mode not a good measure of central tendency?**
The mode is the most commonly observed data item in the data set

**Percentage**

Includes: **Definition, What does Percentage tell you?, How do you calculate the Percentage?, Benefits and downfalls of Percentage**
Percentage is a term used to express a number as a fraction of 100.

**Population, Sample and Estimate**

Includes: **Definitions, What do populations and samples tell us?, Which to use in a survey - Population or Sample?**
A population is any entire group with at least one common characteristic.

**Probability**

Includes: **Definition, What can Probability tell you?, How do you calculate Probability?, Benefits and downfalls of Probability**
Probability refers to the likelihood (or chance) of an event occurring.

**Quantitative and Qualitative**

Includes: **Definitions, Example**
Qualitative data describes an object's attributes that cannot be measured with numbers. Quantitative data is data characterised by numbers.

**Range**

Includes: **Definition, What does the Range tell you?, How is the Range calculated?, Benefits and downfalls of using the Range**
Range represents the distance between the highest and lowest values of a data set.

**Rate**

Includes: **Definition:, What does Rate tell you?, How is the Rate calculated?**
Rate is an expression of ratio; it represents the relationship between two numbers.

**Ratio**

Includes: **Definition, What does Ratio tell you?, How do you calculate the Ratio?**
Ratio is a way of concisely showing the relationship of one quality relative to another.

**Time Series**

Includes: **Definition- Time series, Why are Time-series Created?, Definition- Seasonally Adjusted Time-series, Why are Time-series Seasonally Adjusted?, Definition- Trend Time-series, Why are Trend Time-series Created?, Take care when using Time series**
A time series is a collection of observations obtained through repeated measurements through time.

In this issue

NOTES

STATISTICAL LANGUAGE!

Statistical Language! is an educational resource from the Australian Bureau of Statistics designed to improve the reader's understanding of some fundamental statistical concepts. It is written in plain English for adults and aims to provide them with the basic statistical literacy skills to:

- understand key statistical terminology;
- facilitate access to the expanding level of statistical information presented to the public;
- gain confidence with interpreting summarised information;
- appreciate the importance of statistical information in today's society; and,
- make critical and informed use of data, whatever its source.

All these goals are at the heart of the ABS mission to assist informed decision-making in the Australian community.

Along with simple descriptions, this e-Magazine contains examples and diagrams to help users establish a basic understanding of the key statistical topics covered.

INQUIRIES

For further information about these and related statistics, contact the National Information and Referral Service on 1300 135 070.

About this Release

A glossary of basic statistical terms aimed at helping people understand statistics.

Statistical Language!

CONTENTS

Data

Definitions

Example

Index

Definition

What do indexes tell you?

How can we calculate an Index?

When is an index not appropriate?

Mean

Definition

Here's another way of looking at it!

How do you calculate the mean?

When is the Mean not a useful measure?

Measures of Error and Spread

Definitions

Calculations

Using Standard Deviation, Standard Error and Relative Standard Error

Median

Definition

What can the Median tell you?

How do you calculate a Median?

When is the Median not a good measure of central tendency?

Mode

Definition

What does the mode tell you?

How do you calculate a mode?

When is the mode not a good measure of central tendency?

Percentage

Definition

What does Percentage tell you?

How do you calculate the Percentage?

Benefits and downfalls of Percentage

Population, Sample and Estimate

Definitions

What do populations and samples tell us?

Which to use in a survey - Population or Sample?

Probability

Definition

What can Probability tell you?

How do you calculate Probability?

Benefits and downfalls of Probability

Quantitative and Qualitative

Definitions

Example

Range

Definition

What does the Range tell you?

How is the Range calculated?

Benefits and downfalls of using the Range

Rate

Definition

What does Rate tell you?

How is the Rate calculated?

Ratio

Definition

What does Ratio tell you?

How do you calculate the Ratio?

Time Series

Definition- Time series

Why are Time-series Created?

Definition- Seasonally Adjusted Time-series

Why are Time-series Seasonally Adjusted?

Definition- Trend Time-series

Why are Trend Time-series Created?

Take care when using Time series

Data

Contents >> Data



DATA

This section contains the following subsection :

Definitions
Example

[Previous Page](#)

[Next Page](#)

Definitions

Contents >> Data >> Definitions

DEFINITIONS

Data are observations or facts which once collected, organised and evaluated, become information or knowledge. Data can take various forms, but is often numerical.

A **data item** is the smallest piece of information that can be obtained from a survey.

A **variable** is something measurable that is expected to either change over time or between individual observations. A variable can have any number of data items that relate to a specific measurement.

A **data set** is a collection of data items relating to a number of chosen variables.

[Previous Page](#)

[Next Page](#)

Example

Contents >> Data >> Example

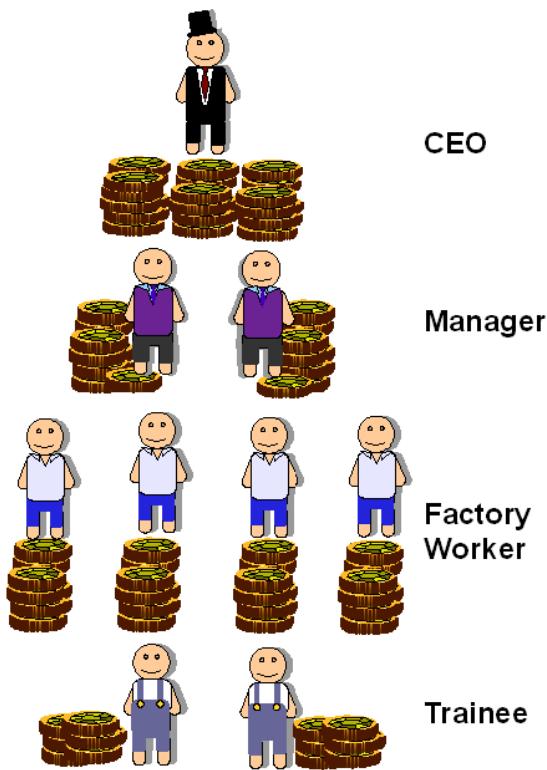
EXAMPLE

Employment details of 'The Widget Company'

Observation	Position in company	Salary (\$)
1	CEO	100 000
2	Manager	60 000
3	Manager	60 000
4	Factory Worker	50 000
5	Factory Worker	50 000
6	Factory Worker	50 000
7	Factory Worker	50 000
8	Trainee	40 000
9	Trainee	40 000

In this table one data item is one salary, for example the CEO salary of \$100,000. The variables are "Position in the Company" and "Salary", and the data set is the entire table.

Data can be portrayed in ways other than tables. For example, the data in the table above is illustrated graphically in the diagram below.



[Previous Page](#)

[Next Page](#)

Index

[Contents >> Index](#)



INDEX

This section contains the following subsection :

[Definition](#)

[What do indexes tell you?](#)

[How can we calculate an Index?](#)

[When is an index not appropriate?](#)

[Previous Page](#)

[Next Page](#)

Definition

[Contents >> Index >> Definition](#)

DEFINITION:

An **Index** is a number used to show the variation in some quantity over time. It is usual to fix the first observation (sometimes called a benchmark) to a base value of 100, then having all the following observations linked to this base to compare any relative changes over time. It is a type of time series data.

[Previous Page](#)

[Next Page](#)

What do indexes tell you?

[Contents >> Index >> What do indexes tell you?](#)

WHAT DO INDEXES TELL YOU?

Indexes are commonly used to track the changes in business and economic conditions. They measure the growth of prices, production and other quantities of economic interest. One of the most widely known index is the Consumer Price Index, which provides a general measure of change in the price of consumer goods and services.

[Previous Page](#)

[Next Page](#)

How can we calculate an Index?

[Contents >> Index >> How can we calculate an Index?](#)

HOW CAN WE CALCULATE AN INDEX?

Indexes are calculated by comparing variations in quantity from one time period to the next. For a price index the calculation is the price from one time period divided by the base period price. In the following example the 'Widget Price Index' started in 2005. The following table demonstrates how a 'Widget Price Index' would be calculated in 2006 and 2007. The percentage change in the price index from 2006 to 2007 is

$$\frac{(124-116)}{116} \times 100\% = 6.9\%$$

Prices of Widgets 2005-2007

Year	Price per Widget (\$)	Widget Price Index
2005	50	100
2006	58	58/50x100=116
2007	62	62/50x100=124

[Previous Page](#)

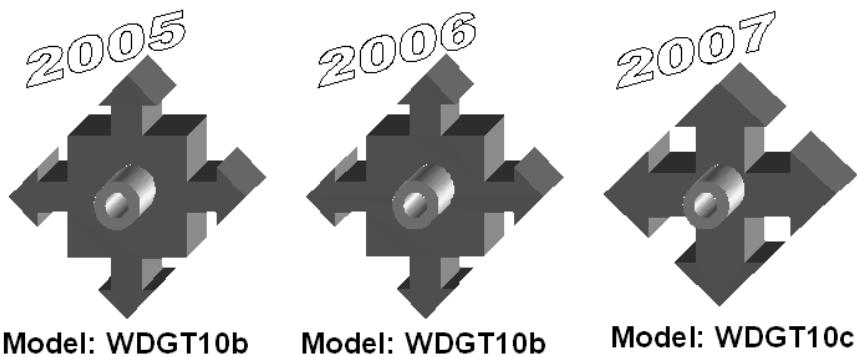
[Next Page](#)

When is an index not appropriate?

[Contents >> Index >> When is an index not appropriate?](#)

WHEN IS AN INDEX NOT APPROPRIATE?

Care should be taken when using indexes. Of particular importance is to ensure the items being monitored do not somehow change over time. For instance, the production cost of an item may not have changed over time, yet in reality the size or the quality of the item may have decreased. For example, the diagram below illustrates the widget associated with the price index.



However it can be seen that in 2007 the widget produced was different from previous years, meaning the index value for 2007 is not necessarily comparable with those of the previous years.

[Previous Page](#)

[Next Page](#)

Mean

[Contents >> Mean](#)



MEAN

This section contains the following subsection :

Definition

Here's another way of looking at it!

How do you calculate the mean?

When is the Mean not a useful measure?

[Previous Page](#)

[Next Page](#)

Definition

[Contents >> Mean >> Definition](#)

DEFINITION

The **mean** is a summary number that measures one type of midpoint in a range of numbers. In statistical terms determining the midpoint in a range of numbers is called the Measure of Central Tendency.

To find the mean of a set of numbers, or observations, we take the total value of all the members of the set and divide it by the number of items in the set. It is also known as the arithmetic average.

[Previous Page](#)

[Next Page](#)

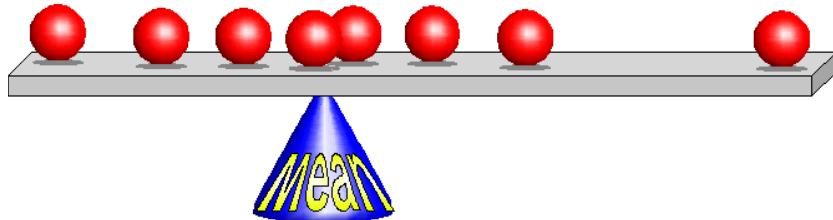
Here's another way of looking at it!

[Contents >> Mean >> Here's another way of looking at it!](#)

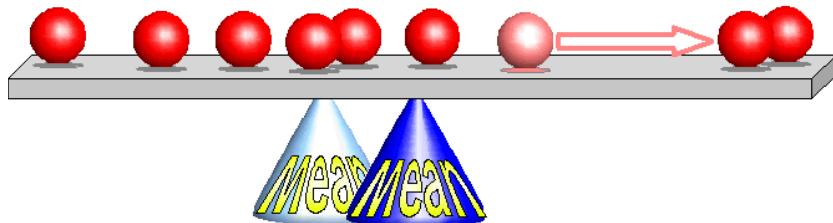
HERE'S ANOTHER WAY OF LOOKING AT IT!

The mean represents the gravitational midpoint of a list of numbers. It can be visualised as a seesaw. A single child can sit farther from the centre of a seesaw in order to balance several children sitting closer to the centre. Similarly, a single value sitting far off the gravitational centre can have a profound effect on the mean.

The mean is the point or value at which the group of observations would balance:



If an observed value in a dataset shifts to the right, the mean will adjust to the new point of balance by also shifting to the right.



[Previous Page](#)

[Next Page](#)

How do you calculate the mean?

[Contents >> Mean >> How do you calculate the mean?](#)

HOW DO YOU CALCULATE THE MEAN?

To calculate the mean, add all the values of observations in the set of numbers and divide by the number of observations.

For example let's look at the salaries of nine people in a company.

Employment details of 'The Widget Company'

Observation	Position in company	Salary (\$)
1	CEO	100 000
2	Manager	60 000
3	Manager	60 000
4	Factory Worker	50 000
5	Factory Worker	50 000
6	Factory Worker	50 000
7	Factory Worker	50 000
8	Trainee	40 000
9	Trainee	40 000

To obtain the mean salary of all employees we do the following:

1. Add together all the salaries.
2. Divide by the number of employees.
3. The result equals the mean.

This is how it will look for the company above:

$$\frac{100,000 + 60,000 + 60,000 + 50,000 + 50,000 + 50,000 + 50,000 + 40,000 + 40,000}{9} = \$55,555.56$$

Thus, the mean income for these nine people is \$55,555.56 - in this case a good indication of the central value of the salaries.

[Previous Page](#)

[Next Page](#)

When is the Mean not a useful measure?

[Contents](#) >> [Mean](#) >> When is the Mean not a useful measure?

WHEN IS THE MEAN NOT A USEFUL MEASURE?

The mean is a good choice of measure of central tendency when the data is more or less symmetrically spread out from the lowest to highest values. However, the mean is not a good measure when the data is unevenly spread.

Let us take the case above, but this time change the CEO's earnings to \$200,000.

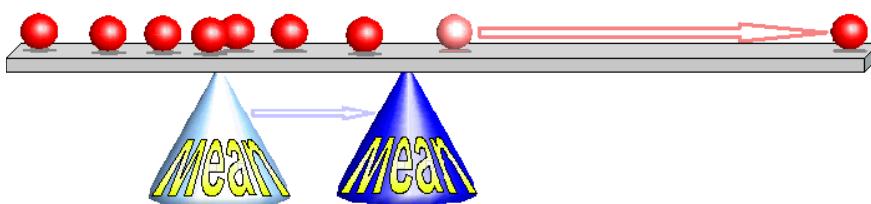
Using the formula again:

$$\frac{200,000 + 60,000 + 60,000 + 50,000 + 50,000 + 50,000 + 50,000 + 40,000 + 40,000}{9} = \$66,666.67$$

The mean income of these nine employees has increased to \$66,666.

Given that only one of the nine employees earns above this amount, the mean is not a good measure of where the midpoint lies.

This is demonstrated in the diagram below.



In this case a much better measure of the midpoint would be the [median](#).

[Previous Page](#)

[Next Page](#)

Measures of Error and Spread

[Contents](#) >> [Measures of Error and Spread](#)



MEASURES OF ERROR AND SPREAD

This section contains the following subsection :

- Definitions
- Calculations
- Using Standard Deviation, Standard Error and Relative Standard Error

[Previous Page](#)

[Next Page](#)

Definitions

[Contents >> Measures of Error and Spread >> Definitions](#)

DEFINITIONS

Standard Deviation:

Standard deviation measures the scatter in a group of observations. It is a calculated summary of the distance each observation in a data set is from the mean. **Standard deviation** gives us a good idea whether a set of observations are loosely or tightly clustered around the mean.

Sampling Error:

Sampling error is the difference between a population characteristic and the estimate of this characteristic based on a sample. Sampling error arises because it's often not possible to collect data on a whole population, and samples aren't often identical in character to their parent population. The larger a sample becomes the more likely it will look like the whole population it was sampled from, resulting in a smaller sampling error. If we did a complete enumeration of the population, such as in a census, there would be no sampling error.

Standard Error:

The **Standard Error** (SE) is one way of measuring the sampling error of an estimate. The theory shows that there are about two chances in three that an estimate from a sample is within one standard error of the true value (the value for the whole population). As such, the larger the standard error, the less confident we are that the estimate from the sample is close to the true value.

There are several types of Standard Error (SE). A commonly used type of standard error in the Australian Bureau of Statistics is the Standard Error of the Mean.

Relative Standard Error:

The relative standard error (RSE) is the standard error of the estimate divided by the estimate itself. It is another way of expressing the standard error to make interpretation easier. It's useful for comparing the size of the standard error across different samples, and is often expressed as a percentage. As with the standard error, the higher the RSE, the less confident we are that the estimate from the sample is close to the true value.

[Previous Page](#)

[Next Page](#)

Calculations

[Contents >> Measures of Error and Spread >> Calculations](#)

CALCULATIONS

Standard Deviation

The following example is a calculation of the standard deviation of the salaries of persons employed in The Widget Company.

Employment details of 'The Widget Company'

Observation	Position in company	Salary (\$)
1	CEO	100 000
2	Manager	60 000
3	Manager	60 000
4	Factory Worker	50 000
5	Factory Worker	50 000
6	Factory Worker	50 000
7	Factory Worker	50 000
8	Trainee	40 000
9	Trainee	40 000

The are six steps to calculating the standard deviation:

1. Calculate the mean

The mean of the salaries in the above table is \$55,555

2. Subtract the mean from each observation

This has been calculated in the above table under the heading 'Difference'. For example, the difference from the mean for the CEO's salary is 100,000 minus 55,555 which equals 44,444.

3. Square each result

This has been calculated in the above table under the heading 'Difference Squared'. For example the difference squared for the CEO's salary is 44,444 times 44,444 which equals 1,975,308,247.

4. Add these squares

In the example this is $1,975,308,246 + 19,753,047 + 19,753,047 + 30,864,247 + 30,864,247 + 30,864,247 + 30,864,247 + 241,975,447 + 241,975,447 = 2,622,222,222$.

5. Divide this sum by the number of observations (the result at this stage is called the "variance").

In this example this is 2,622,222,222 divided by nine (number of observations) which equals 291,358,024.69.

6. Take the positive square root

In this example the square root of 291,358,024.69 is 17,069.21. Therefore, the standard deviation is 17,069.21.

Standard Error

The standard error of the mean of a sample from a population is the standard deviation of the sample divided by the square root of the sample size.

Relative Standard Error

The RSE is generally calculated by with the following formula.

$$RSE = \left(\frac{\text{Standard Error of the Estimate}}{\text{Estimate value}} \right) \times 100\%$$

[Previous Page](#)

[Next Page](#)

Using Standard Deviation, Standard Error and Relative Standard Error

[Contents](#) >> [Measures of Error and Spread](#) >> [Using Standard Deviation, Standard Error and Relative Standard Error](#)

USING STANDARD DEVIATION, STANDARD ERROR AND RELATIVE STANDARD ERROR

Because of its close links with the mean, the standard deviation can be seriously affected if the mean is a poor measure of central location. The standard deviation is also very sensitive to observations well outside where the majority of observations are found (outliers); therefore, it is most useful in regularly distributed sets of data.

The Relative Standard Error is used as an indication of the reliability of the data. Let us look at the data below extracted from the Survey of Persons with Two Cars.

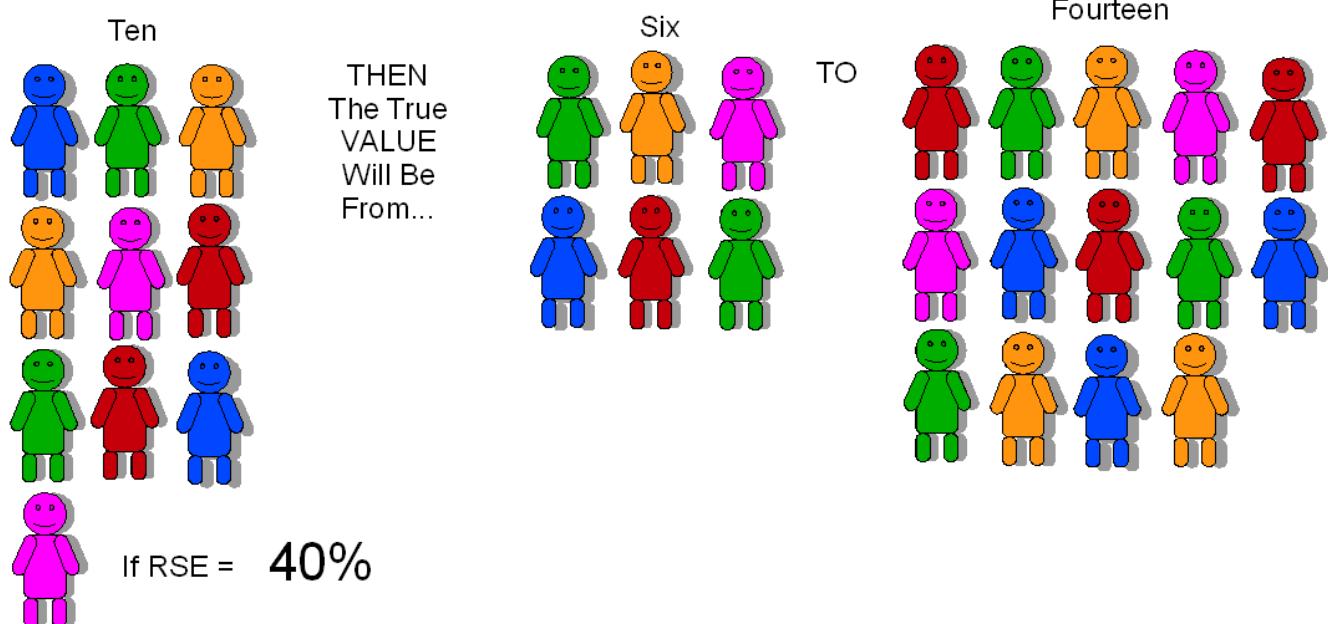
survey of number of persons with 2 cars

	persons with two cars	RSE (%)
persons with 2 cars	10	40
persons without 2 cars	50	5

In the case above a RSE of 40% for 10 people who have two cars means that the true value could be four persons either side of 10 persons. So the true value is likely to be between six and 14 persons.

Estimate

Real Value



[Previous Page](#)

[Next Page](#)

Median

[Contents >> Median](#)



MEDIAN

This section contains the following subsection :

Definition

What can the Median tell you?

How do you calculate a Median?

When is the Median not a good measure of central tendency?

[Previous Page](#)

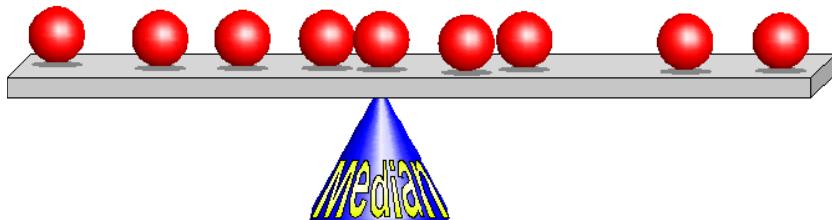
[Next Page](#)

Definition

[Contents >> Median >> Definition](#)

DEFINITION

Median is one of the three measures of central tendency. A median is the middle score that separates the higher half of a data set from the lower half. It looks at the midpoint of a set of data when the numbers are ordered numerically.



What can the Median tell you?

Contents >> Median >> What can the Median tell you?

WHAT CAN THE MEDIAN TELL YOU?

The median provides a helpful measure of the centre of a dataset. By comparing the median to the mean, you can get an idea of the distribution of a dataset. When the mean and the median are the same, the dataset is more or less evenly distributed from the lowest to highest values. When the mean and the median are different then it is likely the data is not symmetrical but is either skewed to the left or the right.

How do you calculate a Median?

Contents >> Median >> How do you calculate a Median?

HOW DO YOU CALCULATE A MEDIAN?

To calculate the median, line up all the values in the data set from smallest to largest and identify the middle value. If there are even numbers of values, then the median is the average of the middle two values.

For example, let's look at 'The Widget Company' and the salaries of its nine employees.

Employment details of 'The Widget Company'

Observation	Position in company	Salary (\$)
1	CEO	100 000
2	Manager	60 000
3	Manager	60 000
4	Factory Worker	50 000
5	Factory Worker	50 000
6	Factory Worker	50 000
7	Factory Worker	50 000
8	Trainee	40 000
9	Trainee	40 000

To identify which value in a data set is the median we use the following formula:

$$\text{Median} = \frac{(\text{Number of Observations} + 1)}{2} = \frac{(9 + 1)}{2} = 5$$

The fifth observation of the data set (when the data set is arranged from lowest to highest) will have the median salary. So the median salary of workers at the 'The Widget Company' is \$50,000, a good indication of a central value of the salaries.



When is the Median not a good measure of central tendency?

Contents >> Median >> When is the Median not a good measure of central tendency?

WHEN IS THE MEDIAN NOT A GOOD MEASURE OF CENTRAL TENDENCY?

The median can be a good measure of central tendency, especially when the data is evenly distributed. It can illustrate a skewed and uneven distribution when compared to the mean. However, there are occasions when the median is not always a good measure of central tendency. As the median only reports on the middle most observation it does ignore the actual values of all the rest of the data, one outcome is that the median may hide the presence of extreme values.

Let us consider the wages of 'The Widget Company', we will increase the earnings of the CEO from \$100,000 to \$500,000.

When the CEO's salary changes to \$500,000 we can see that the median income remains at \$50,000, even though the CEO now earns five times their previous income.



[Previous Page](#)

[Next Page](#)

Mode

[Contents >> Mode](#)



MODE

This section contains the following subsection :

[Definition](#)

[What does the mode tell you?](#)

[How do you calculate a mode?](#)

[When is the mode not a good measure of central tendency?](#)

[Previous Page](#)

[Next Page](#)

Definition

[Contents >> Mode >> Definition](#)

DEFINITION

The mode is one of the three measures of central tendency. It is the most commonly observed data item in a data set.

[Previous Page](#)

[Next Page](#)

What does the mode tell you?

[Contents >> Mode >> What does the mode tell you?](#)

WHAT DOES THE MODE TELL YOU?

The mode gives us the observation most likely to occur.

[Previous Page](#)

[Next Page](#)

How do you calculate a mode?

Contents >> Mode >> How do you calculate a mode?

HOW DO YOU CALCULATE A MODE?

To calculate the mode, locate the most commonly occurring data item.

Example

Employment details of 'The Widget Company'

Observation	Position in company	Salary (\$)
1	CEO	100 000
2	Manager	60 000
3	Manager	60 000
4	Factory Worker	50 000
5	Factory Worker	50 000
6	Factory Worker	50 000
7	Factory Worker	50 000
8	Trainee	40 000
9	Trainee	40 000

To obtain the mode we identify the most common salary. In the case of the above table the mode is \$50,000.



[Previous Page](#)

[Next Page](#)

When is the mode not a good measure of central tendency?

Contents >> Mode >> When is the mode not a good measure of central tendency?

WHEN IS THE MODE NOT A GOOD MEASURE OF CENTRAL TENDENCY?

There are a number of reasons when the mode would not be the best measure of central tendency. Some are illustrated below.

Example

If there is more than one mode these modes will not be meaningful if a single measure of central tendency is required.

Employment details of 'The Widget Company'

Observation	Position in company	Salary (\$)
1	CEO	100 000
2	Manager	100 000
3	Manager	59 000
4	Factory Worker	52 000
5	Factory Worker	51 000
6	Factory Worker	50 000
7	Factory Worker	49 000
8	Trainee	40 000
9	Trainee	40 000

In the above table there are two modes, 100,000 and 40,000, therefore it is impossible to determine one measure of central tendency using mode in this case.



Example

If there is no mode in a data set (as all data items are different) then a different measure of central tendency is required.

Employment details of 'The Widget Company'

Observation	Position in company	Salary (\$)
1	CEO	100 000
2	Manager	98 000
3	Manager	59 000
4	Factory Worker	52 000
5	Factory Worker	51 000
6	Factory Worker	50 000
7	Factory Worker	49 000
8	Trainee	45 000
9	Trainee	40 000

In the above table there is no mode as all salaries are different.



Example

The mode may not fall near the centre of the data.

Employment details of 'The Widget Company'

Observation	Position in company	Salary (\$)
1	CEO	100 000
2	Manager	98 000
3	Manager	59 000
4	Factory Worker	52 000
5	Factory Worker	51 000
6	Factory Worker	50 000
7	Factory Worker	49 000
8	Trainee	40 000
9	Trainee	40 000

In the above table the mode is now \$40,000, which is not a good measure of central tendency.



Percentage

[Contents >> Percentage](#)

PERCENTAGE

This section contains the following subsection :

[Definition](#)[What does Percentage tell you?](#)[How do you calculate the Percentage?](#)[Benefits and downfalls of Percentage](#)

Definition

[Contents >> Percentage >> Definition](#)

DEFINITION

Percentage is the term used to express a number as a fraction of one hundred, it compares one value in relation to another. A percentage is often symbolised using the percent sign %.

What does Percentage tell you?

[Contents >> Percentage >> What does Percentage tell you?](#)

WHAT DOES PERCENTAGE TELL YOU?

Percentage is commonly used to represent statistical data; it is considered an important tool to illustrate the proportion of something. The percentage total of a data set should always add up to 100 except in special circumstances. Percentages larger than the value of 100 often occur in financial situations, say for instance if a item originally costing \$1 was sold for \$1 then the profit would be 0%, if the same item was sold for \$2 then the profit would be 100%, and selling it for \$3 would be a 200% profit.

How do you calculate the Percentage?

[Contents >> Percentage >> How do you calculate the Percentage?](#)

HOW DO YOU CALCULATE THE PERCENTAGE?

The percentage of a data set can be calculated by dividing a component value by the total value; and then multiplying that value by 100.

$$\text{Percentage \%} = \left(\frac{\text{Value}}{\text{Total}} \right) \times 100\%$$

Example

If we wanted to know the percentage of employees who are factory workers out of all employees, the same formula is applied.

Employment details of 'The Widget Company'

Observation	Position in company	Salary (\$)
1	CEO	100 000
2	Manager	60 000
3	Manager	60 000
4	Factory Worker	50 000
5	Factory Worker	50 000

In this case:

$$\text{Factory Workers \%} = \left(\frac{\text{Number of Factory Workers}}{\text{Total number of Employees}} \right) \times 100\% = \left(\frac{4}{9} \right) \times 100\% = 44\%$$

Thus we find 44% of all employees from 'The Widget Company' are factory workers.

[Previous Page](#)

[Next Page](#)

Benefits and downfalls of Percentage

[Contents](#) >> [Percentage](#) >> Benefits and downfalls of Percentage

BENEFITS AND DOWNFALLS OF PERCENTAGE

Percentage represents the proportions of known values. By using a percentage to represent the data you provide the reader with an easy to read summary statistic. There can be downfalls of using the percentage. Percentages are not additive, as each percentage calculation is worked out on the new value. For instance, and 10% pay rise followed by a 10% wage cut results in an overall pay decrease, since the wage cut is based on a larger number than the original number the wage rise was based on.



Similarly, a shop giving 50% off all stock with a further 50% off selected items does not mean the selected items are free.



[Previous Page](#)

[Next Page](#)

Population, Sample and Estimate

[Contents](#) >> [Population, Sample and Estimate](#)



POPULATION, SAMPLE AND ESTIMATE

This section contains the following subsection :

Definitions

What do populations and samples tell us?

Which to use in a survey - Population or Sample?

[Previous Page](#)

[Next Page](#)

Definitions

DEFINITIONS

Population

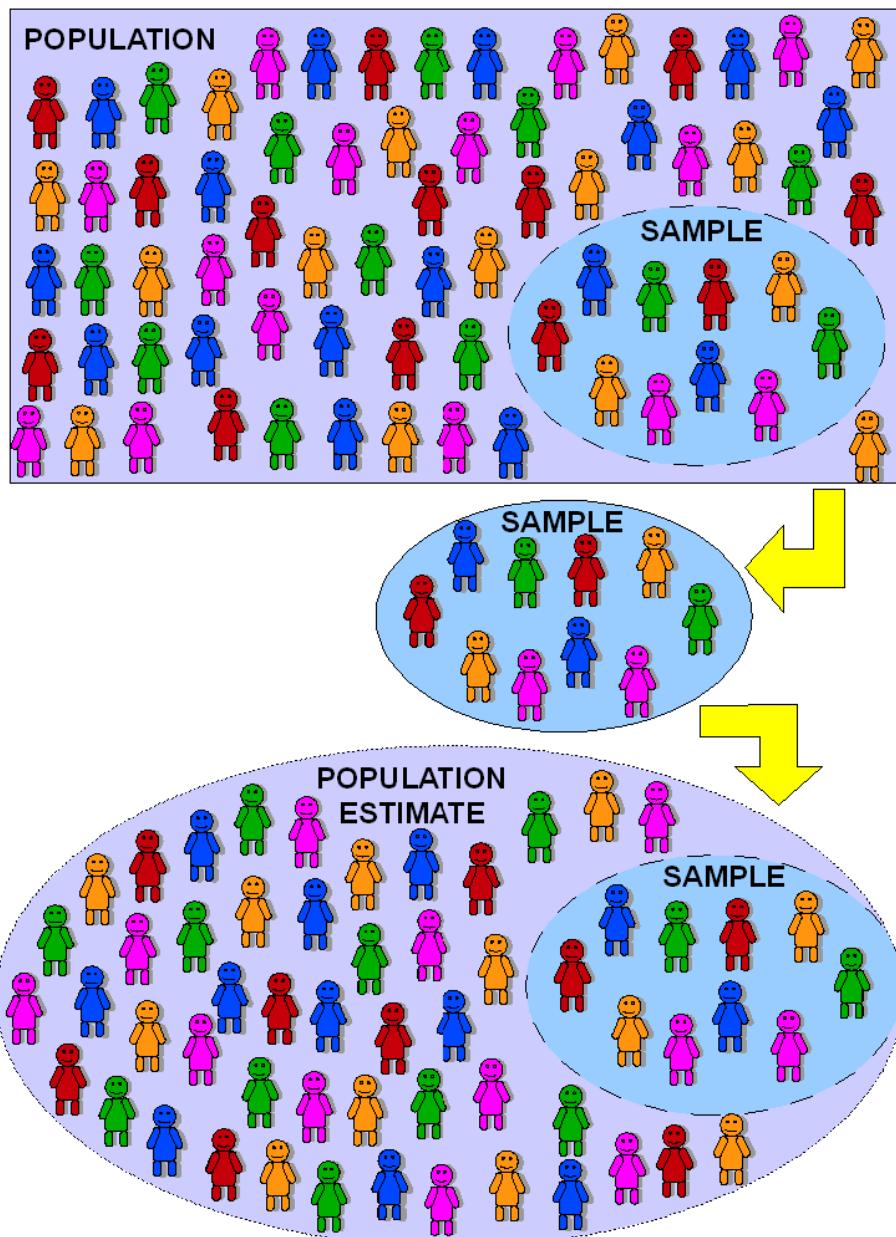
A population is any entire group with at least one characteristic in common. A survey covering the a population is commonly called a census. In a census every member of the group is included and the entire group is used to characterise the population.

Sample

A sample is part of a population. It is a subset of the population, often randomly selected for the purpose of studying the characteristics of the entire population.

Estimate

An estimate is information about a population extrapolated from a sample of the population.



What do populations and samples tell us?

WHAT DO POPULATIONS AND SAMPLES TELL US?

Ideally, an investigator would prefer to get the characteristics of a whole population, but most times this is not possible. In order to make any generalisations about a population, a sample is used as a representative of the population. However, for any population there are many possible samples.

Example

The population for a study of infant health might be all children born in Australia from 1980 to 1989. The sample might be all babies born on 7th May in any of those years.

[Previous Page](#)

[Next Page](#)

Which to use in a survey - Population or Sample?

[Contents](#) >> [Population, Sample and Estimate](#) >> [Which to use in a survey - Population or Sample?](#)

WHICH TO USE IN A SURVEY - POPULATION OR SAMPLE?

Many factors must be considered when developing a survey. In most situations the best statistics come from using the whole population. But for various reasons this is not always practical. The type of information obtained will be affected by the type of survey conducted and data collected. Some of these are listed below.

Time - It takes time to run a survey, surveying an entire population can be very long.

Expense - Conducting a survey for an entire population can be expensive for large populations.

Accuracy - The larger the sample the more likely it will display the characteristics of the population.

Information - A large sample can be further broken down to smaller subgroups for detailed analysis.

Confidentiality - When personal information is collected in the ABS surveys must be large enough so that no individual can be identified.

[Previous Page](#)

[Next Page](#)

Probability

[Contents](#) >> [Probability](#)



PROBABILITY

This section contains the following subsection :

Definition

What can Probability tell you?

How do you calculate Probability?

Benefits and downfalls of Probability

[Previous Page](#)

[Next Page](#)

Definition

[Contents](#) >> [Probability](#) >> [Definition](#)

DEFINITION

Probability refers to the likelihood (or chance) of an event occurring. The term itself is extensively used in statistics because it can be used to calculate an expected frequency.

[Previous Page](#)

[Next Page](#)

What can Probability tell you?

Contents >> Probability >> What can Probability tell you?

WHAT CAN PROBABILITY TELL YOU?

When an event can be repeated and the results of that event have some underlying pattern then probability can be used to describe that pattern and how often a result could occur. It is expressed as a value between zero and one. When the value is zero, there is no chance the predicted event will occur, when the value is one the event will always occur. As the probability value increases the chance of a predicted event occurring also increases.

Previous Page

Next Page

How do you calculate Probability?

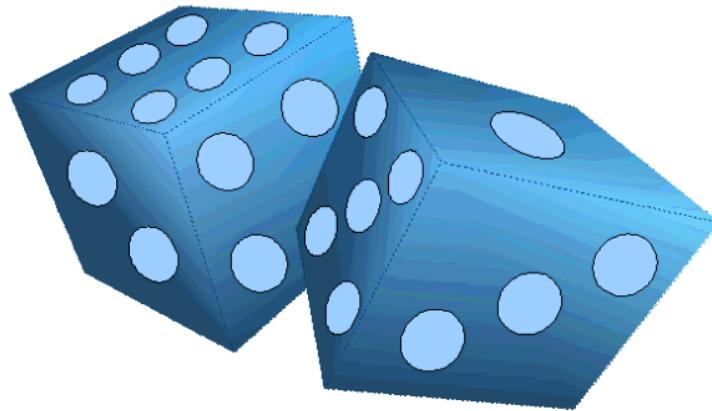
Contents >> Probability >> How do you calculate Probability?

HOW DO YOU CALCULATE PROBABILITY?

Probability is calculated by dividing the number of ways an event can occur by the total number of possible outcomes.

$$\text{Probability} = \frac{\text{Number of Ways an Event can Occur}}{\text{All Possible Outcomes}}$$

For instance, the chance of rolling two 'ones' on a pair of dice is once out of 36 different combinations or about 0.06, to roll the sum of seven would be about 0.17, and the probability of rolling the sum of 13 on a pair of dice would be 0.



Example

To calculate a probability using a working example, let's look at 'The Widget Company'. If we put all the employees names into a hat what is the chance that we'll pull out the name of the CEO?

Employment details of 'The Widget Company'

Observation	Position in company
1	CEO
2	Manager
3	Manager
4	Factory Worker
5	Factory Worker
6	Factory Worker
7	Factory Worker
8	Trainee
9	Trainee

We can see that there are nine employees in total and only one CEO.

$$\text{Probability(CEO)} = \frac{\text{Number of CEOs}}{\text{Number of ALL Employees}} = \frac{1}{9} = 0.11$$

Therefore, there is a one in nine chance of randomly picking the CEO out of the hat. As we said before we can also express this degree of likelihood as a number between one and zero, in this case 0.11.

Previous Page

Next Page

Benefits and downfalls of Probability

BENEFITS AND DOWNFALLS OF PROBABILITY

Probability allows us to predict the likelihood of an event occurring when it is repeated several times. One of the major benefits of probability is that its fraction can be converted into a percentage. This in turn means that useful predictions can be made in a wide range of areas, including economics, population growth, gambling and even the weather.

[Previous Page](#)

[Next Page](#)

Quantitative and Qualitative

Contents >> Quantitative and Qualitative



QUANTITATIVE AND QUALITATIVE

This section contains the following subsection :

- Definitions
- Example

[Previous Page](#)

[Next Page](#)

Definitions

Contents >> Quantitative and Qualitative >> Definitions

DEFINITIONS

Qualitative data is data describing an object's attributes and is distinguished by non-numeric characteristics. Arithmetic operations cannot be applied. Some typical qualitative data types are eye colour, flavour and product type.

Quantitative data is data characterised by numbers. Arithmetic operations can be applied to the data, for example mean and median calculations.

[Previous Page](#)

[Next Page](#)

Example

Contents >> Quantitative and Qualitative >> Example

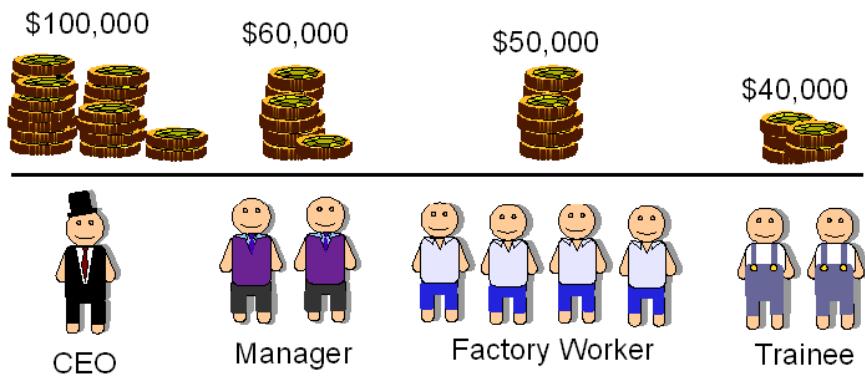
EXAMPLE

In the table employment types and positions in the company are qualitative data and salaries are quantitative data.

Employment details of 'The Widget Company'

Observation	Position in company	Salary (\$)
1	CEO	100 000
2	Manager	60 000
3	Manager	60 000
4	Factory Worker	50 000
5	Factory Worker	50 000
6	Factory Worker	50 000
7	Factory Worker	50 000
8	Trainee	40 000
9	Trainee	40 000

The diagram below illustrates the same qualitative and quantitative data. The data above the line is quantitative, and the data below the line qualitative.



[Previous Page](#)

[Next Page](#)

Range

[Contents >> Range](#)

x-y

RANGE

This section contains the following subsection :

Definition

What does the Range tell you?

How is the Range calculated?

Benefits and downfalls of using the Range

[Previous Page](#)

[Next Page](#)

Definition

[Contents >> Range >> Definition](#)

DEFINITION

Range is the simplest measure of the spread of a data set; it represents the separation distance between the highest and lowest values. The range defines the perimeters of a set of data.

[Previous Page](#)

[Next Page](#)

What does the Range tell you?

[Contents >> Range >> What does the Range tell you?](#)

WHAT DOES THE RANGE TELL YOU?

The range can only tell you basic details about the spread of a set of data. By giving the difference between the lowest and highest scores of a set of data it gives a rough idea of how widely spread out the most extreme observations are, but gives no information as to where any of the other data points lie.

[Previous Page](#)

[Next Page](#)

How is the Range calculated?

[Contents >> Range >> How is the Range calculated?](#)

HOW IS THE RANGE CALCULATED?

The range is calculated by subtracting the lowest value from the highest value of a data set.

Range = Highest Value – Lowest Value

Let us look at the following working example.

Employment details of 'The Widget Company'

Observation	Position in company	Salary (\$)
1	CEO	100 000
2	Manager	60 000
3	Manager	60 000
4	Factory Worker	50 000
5	Factory Worker	50 000
6	Factory Worker	50 000
7	Factory Worker	50 000
8	Trainee	40 000
9	Trainee	40 000

To get the range of the salaries, the lowest value is subtracted from the highest value.

Range of Salaries = \$100,000 – \$40,000 = \$60,000

Therefore, the range of salaries within the company is \$60,000. This illustrates that there is a lot of variability between individual salaries.

[Previous Page](#)

[Next Page](#)

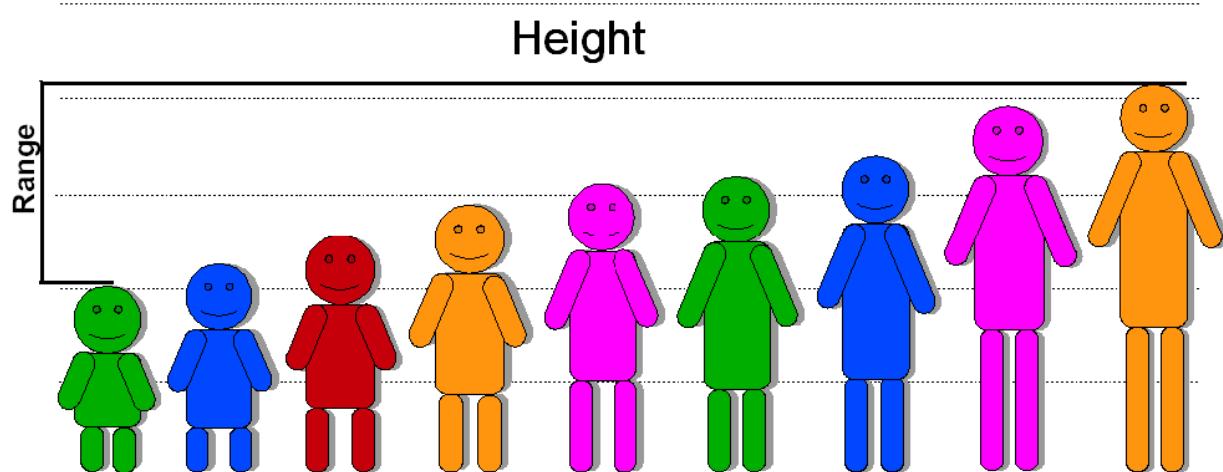
Benefits and downfalls of using the Range

[Contents >> Range >> Benefits and downfalls of using the Range](#)

BENEFITS AND DOWNFALLS OF USING THE RANGE

The range is easy to calculate and understand and is therefore a reasonable measure of spread. However, it should only be used to supplement other measurements of spread as it gives no details of the observations in between.

The following diagram below illustrates the range of heights.



[Previous Page](#)

[Next Page](#)

Rate

[Contents >> Rate](#)



This section contains the following subsection :

Definition

What does Rate tell you?

How is the Rate calculated?

[Previous Page](#)

[Next Page](#)

Definition

[Contents >> Rate >> Definition](#)

DEFINITION:

Rate is an expression of the ratio; it represents the relationship between two measurements. While the division operator of a ratio is a colon, the division operator of a rate is per. The rate is used when the two variables are of different measurements (or units).

[Previous Page](#)

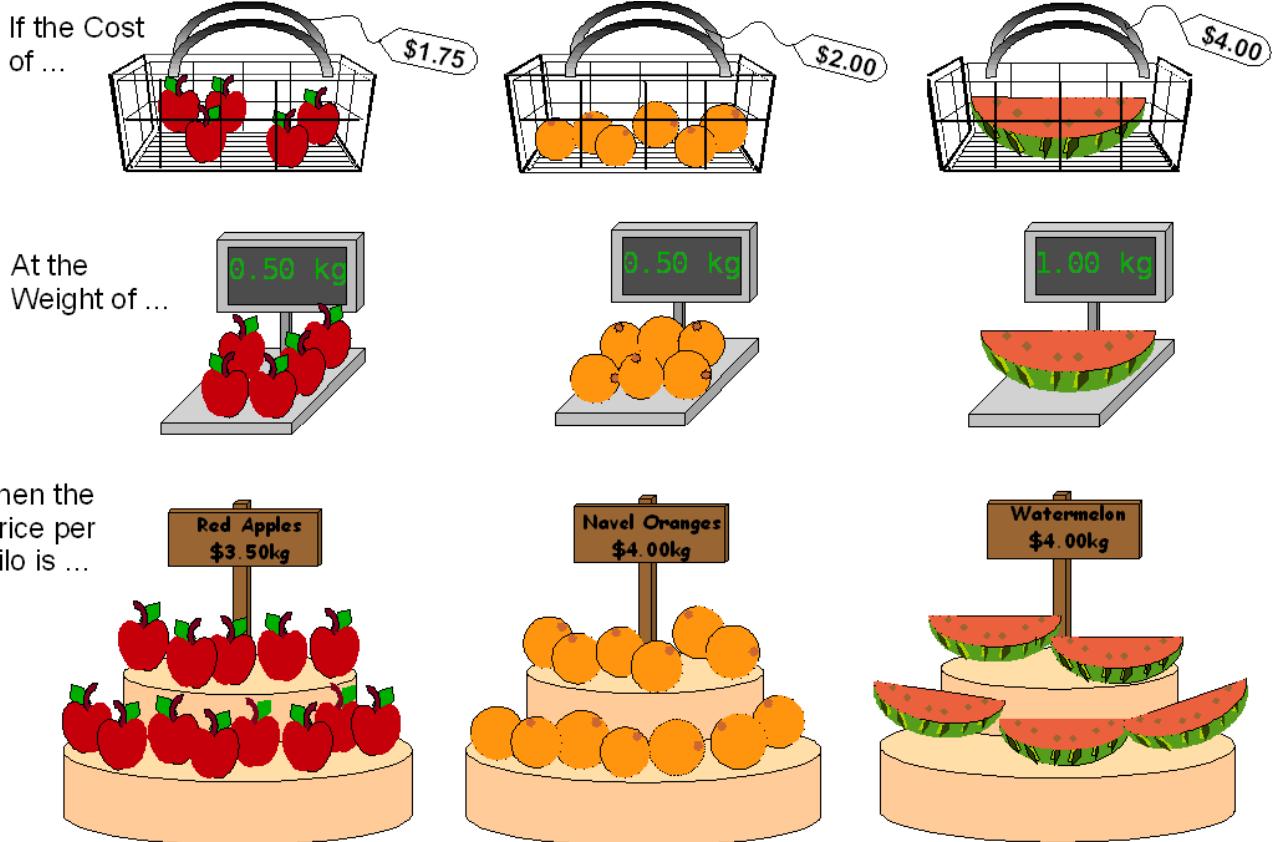
[Next Page](#)

What does Rate tell you?

[Contents >> Rate >> What does Rate tell you?](#)

WHAT DOES RATE TELL YOU?

Rate allows you to compare and define the relationship of two variables which both have its own unit of measurement. For example, when you're buying fruit, the rate used is 'kilos per dollar'.



Other common examples of rate include: cents per litre (price of fuel); pounds per square inch (tyre pressure); and kilometres per hour (speed).

[Previous Page](#)

[Next Page](#)

How is the Rate calculated?

[Contents >> Rate >> How is the Rate calculated?](#)

HOW IS THE RATE CALCULATED?

To calculate the rate of two different units, you divide the units to get a proportion. The denominator is usually the value of most interest. If we return to 'The Widget Company', the rate of money earned can be calculated by dividing the pay by number of hours worked.

If we know a manager receives \$1,150 per week and works thirty hours a week, we can work out the amount of pay for each hour worked.

$$\text{Rate of Pay} = \frac{\$1,150}{30 \text{ hours}} = \$38.33 \text{ per hour}$$

A manager working for 'The Widget Company' receives \$38.33 per hour worked.

[Previous Page](#)

[Next Page](#)

Ratio

[Contents >> Ratio](#)



This section contains the following subsection :

Definition

What does Ratio tell you?

How do you calculate the Ratio?

[Previous Page](#)

[Next Page](#)

Definition

[Contents >> Ratio >> Definition](#)

DEFINITION

A **ratio** is a way of concisely showing the relationship of one quantity relative to another. When the variables are of different measurements we can express the ratio as a **rate**.

[Previous Page](#)

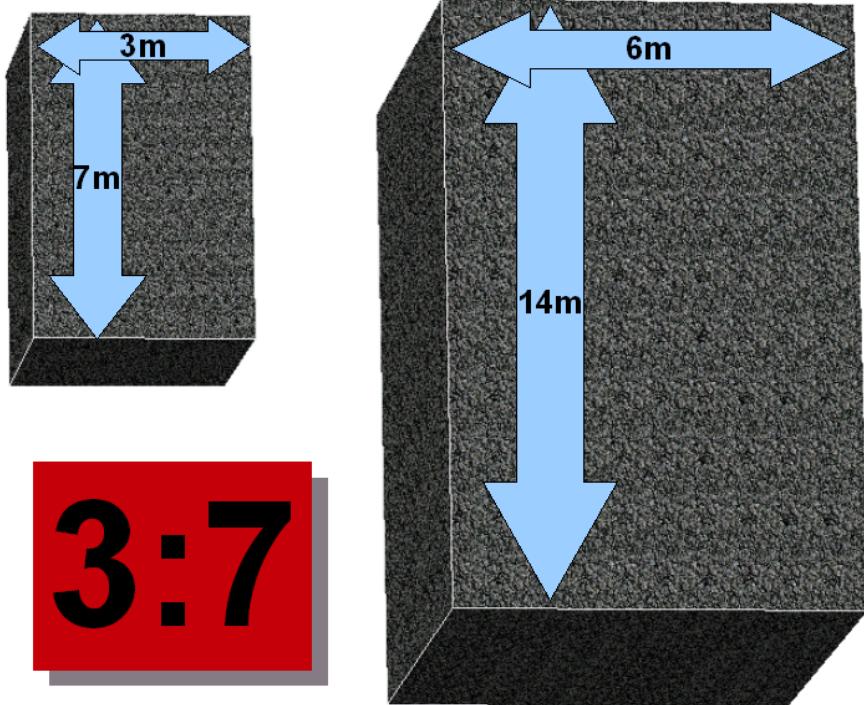
[Next Page](#)

What does Ratio tell you?

[Contents >> Ratio >> What does Ratio tell you?](#)

WHAT DOES RATIO TELL YOU?

A ratio is represented by two whole numbers separated by a colon ' : '. The colon itself is a symbolisation of the word 'to'. For example, 3:7 is read as three to seven. Both shed floors below have the same ratio even though one is 3 metres wide by 7 metres long and the other is 6 metres wide by 14 metres long.



Fractions and percentages are both applications of ratios. Fractions relate the part (the numerator) to the whole (the denominator) while percentages indicate parts per 100.

[Previous Page](#)

[Next Page](#)

How do you calculate the Ratio?

[Contents >> Ratio >> How do you calculate the Ratio?](#)

HOW DO YOU CALCULATE THE RATIO?

Like the percentage, the ratio is obtained by comparing two values of interest. Using our 'The Widget Company' example, we can find the ratio of employees by status.

Employment details of 'The Widget Company'

Observation	Position in company	Salary (\$)
1	CEO	100 000
2	Manager	60 000
3	Manager	60 000
4	Factory Worker	50 000
5	Factory Worker	50 000
6	Factory Worker	50 000
7	Factory Worker	50 000
8	Trainee	40 000
9	Trainee	40 000

To find the ratio of trainees and factory workers from 'The Widget Company', count the number of trainees compare it to the number of factory workers.

Thus the ratio is 2:4 or 1:2.

[Previous Page](#)

[Next Page](#)

Time Series

[Contents >> Time Series](#)



TIME SERIES

This section contains the following subsection :

- Definition- Time series
- Why are Time-series Created?
- Definition- Seasonally Adjusted Time-series
- Why are Time-series Seasonally Adjusted?
- Definition- Trend Time-series
- Why are Trend Time-series Created?
- Take care when using Time series

[Previous Page](#)

[Next Page](#)

Definition- Time series

[Contents >> Time Series >> Definition- Time series](#)

DEFINITION- TIME SERIES

A time-series is a collection of observations obtained through repeated measurements over time. A time-series has three components: the trend (long term direction); the seasonal (systematic, calendar related movements); and, the irregular (unsystematic, short term fluctuations).

[Previous Page](#)

[Next Page](#)

Why are Time-series Created?

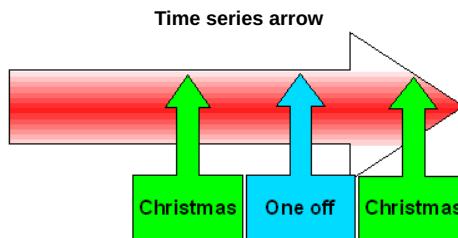
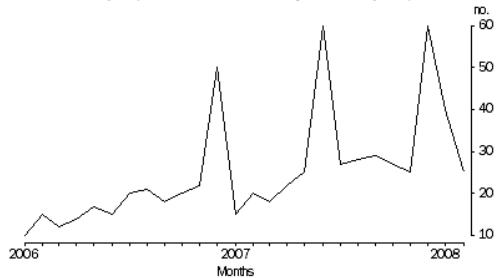
[Contents >> Time Series >> Why are Time-series Created?](#)

WHY ARE TIME-SERIES CREATED?

Collecting observations over time allows us to measure changes over a fixed period, and to compare changes from one time period to another.

Example of an Time-series:

Total Overtime for Employees of 'The Widget Company' 2006-2007, Original



This graphs shows the total amount of overtime worked by employee's each month over a two year period. The series has not been adjusted for seasonal or irregular affects, making it difficult to visualise the underlying trends.

[Previous Page](#)

[Next Page](#)

Definition- Seasonally Adjusted Time-series

[Contents >> Time Series >> Definition- Seasonally Adjusted Time-series](#)

DEFINITION- SEASONALLY ADJUSTED TIME-SERIES

A seasonally adjusted time-series is a time-series with seasonal component removed. This component shows a pattern over one year or less and is systemic or calendar related.

[Previous Page](#)

[Next Page](#)

Why are Time-series Seasonally Adjusted?

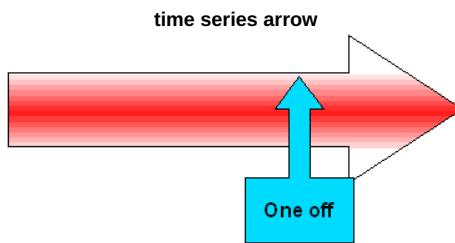
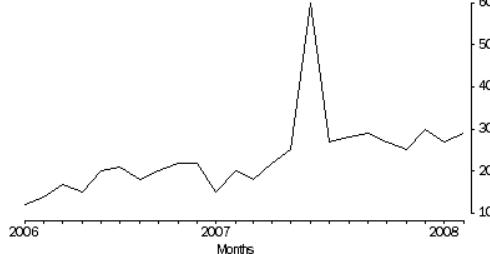
[Contents >> Time Series >> Why are Time-series Seasonally Adjusted?](#)

WHY ARE TIME-SERIES SEASONALLY ADJUSTED?

Time series are seasonally adjusted as seasonal effects can conceal both the true underlying movement in the series, as well as certain non-seasonal characteristics which may be of interest.

Example of a Seasonally Adjusted Time Series:

Total Overtime for Employees of 'The Widget Company' 2006-2007, Seasonally Adjusted



The time-series has been adjusted to remove the sharp increase in the amount of overtime worked in December, which is attributable to the increased demand for Christmas Widgets over the Christmas season. The series has not been adjusted for irregular or one-off affects, which have the potential to mask underlying trends.

[Previous Page](#)

[Next Page](#)

Definition- Trend Time-series

[Contents >> Time Series >> Definition- Trend Time-series](#)

DEFINITION- TREND TIME-SERIES

A trend time-series is a seasonally adjusted time-series that has been further adjusted to remove irregular components. Examples of irregular events include: weather damage to a fruit crop causing a price jump over a short period; a terrorist attack causing a brief drop in tourism; or, an advertising campaign creating a demand in a product while the campaign runs.

[Previous Page](#)

[Next Page](#)

Why are Trend Time-series Created?

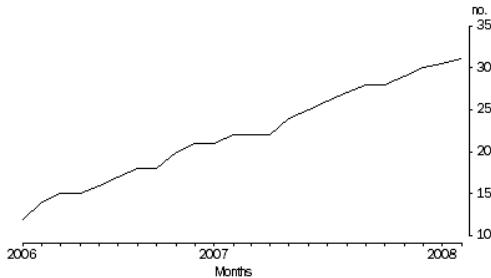
[Contents >> Time Series >> Why are Trend Time-series Created?](#)

WHY ARE TREND TIME-SERIES CREATED?

Trend time-series are created in order to view the long-term trend. Irregular components can be removed in a number of ways, for example by using weighted averages or by applying a filter to the data.

Example of a Trend Time -series:

Total Overtime for Employees of 'The Widget Company' 2006-2007, Trend



Time series arrow 3



The June 2007 figure has been adjusted by the greatest amount because it is an irregular occurrence for a high amount of overtime to be worked during this month. For example, this may represent a one off demand for the installation of new factory equipment. In addition, the other months have been smoothed to reveal the underlying trends.

[Previous Page](#)

[Next Page](#)

Take care when using Time series

[Contents >> Time Series >> Take care when using Time series](#)

TAKE CARE WHEN USING TIME SERIES

Because time series data provides us with information on observations over time, users must be aware of changes of the measured items over time. A similar issue exists with [indexes](#).

[Previous Page](#)

© Commonwealth of Australia

All data and other material produced by the Australian Bureau of Statistics (ABS) constitutes Commonwealth copyright administered by the ABS. The ABS reserves the right to set out the terms and conditions for the use of such material. Unless otherwise noted, all material on this website - except the ABS logo, the Commonwealth Coat of Arms, and any material protected by a trade mark - is licensed under a Creative Commons Attribution 2.5 Australia licence